Noise-Tolerant Life-Long Matrix Completion via Adaptive Sampling

Hongyang Zhang
Joint work with Maria-Florina Balcan
Machine Learning Department, CMU

Machine Learning Lunch

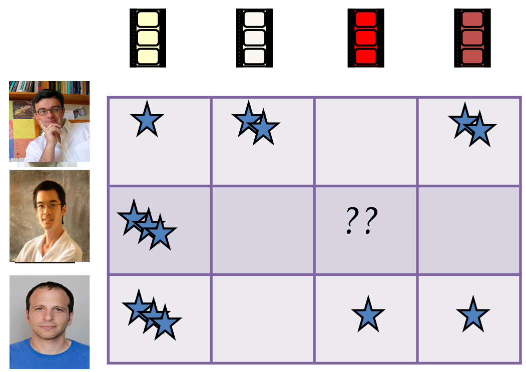
Life-Long Matrix Completion

Real-world **applications:**

Recommendation System

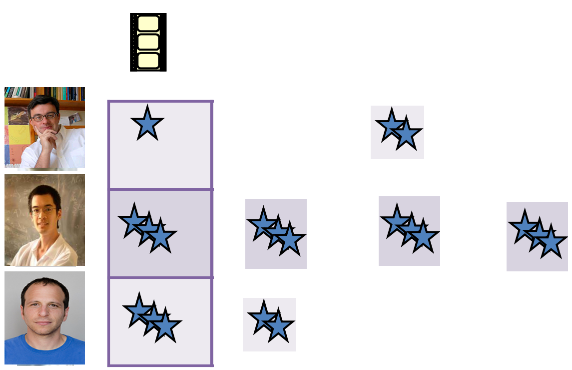
Compressed Sensing

Matrix Completion

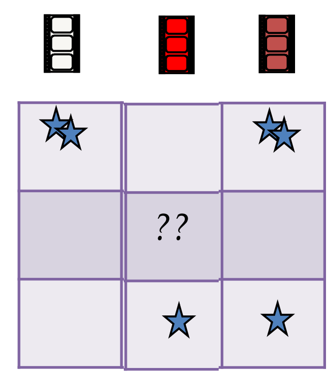


What if the data comes online?

Arrived Columns



Coming Columns

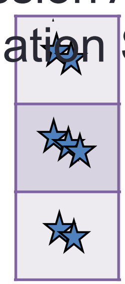


Outline

- Motivation and examples
- **Our goal and approach**
- Matrix completion background
- Robustness Analysis
- Experimental Results

Our Sampling Model

- Adaptive Sampling
  - Scheme 1: Uniformly take the samples randomly (smaller sample complexity)
  - Scheme 2: Request all entries of column from oracle (larger sample complexity)
- Sampling scheme in the real world
  - Network Tomography
  - Gene Expression Analysis
  - Recommendation System



Goal: Keep Sample Complexity as small as possible

Krishnamurthy and Singh, Low-Rank Matrix and Tensor Completion via Adaptive Sampling, NIPS 2013

Our Approach

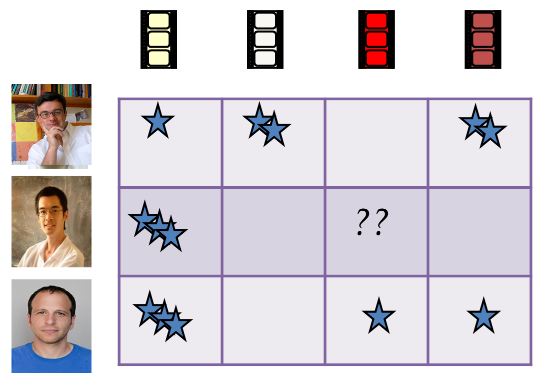
Real-world **applications:**

Recommendation System

Compressed Sensing

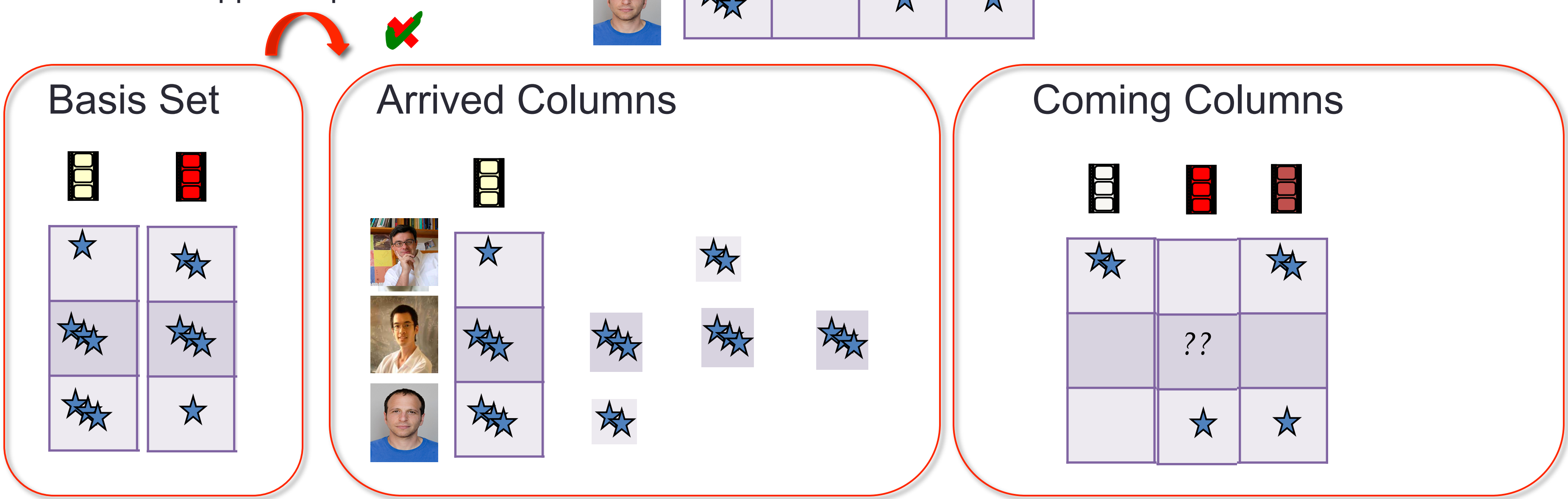
Appro Representable?

Matrix Completion



What if the data comes online?

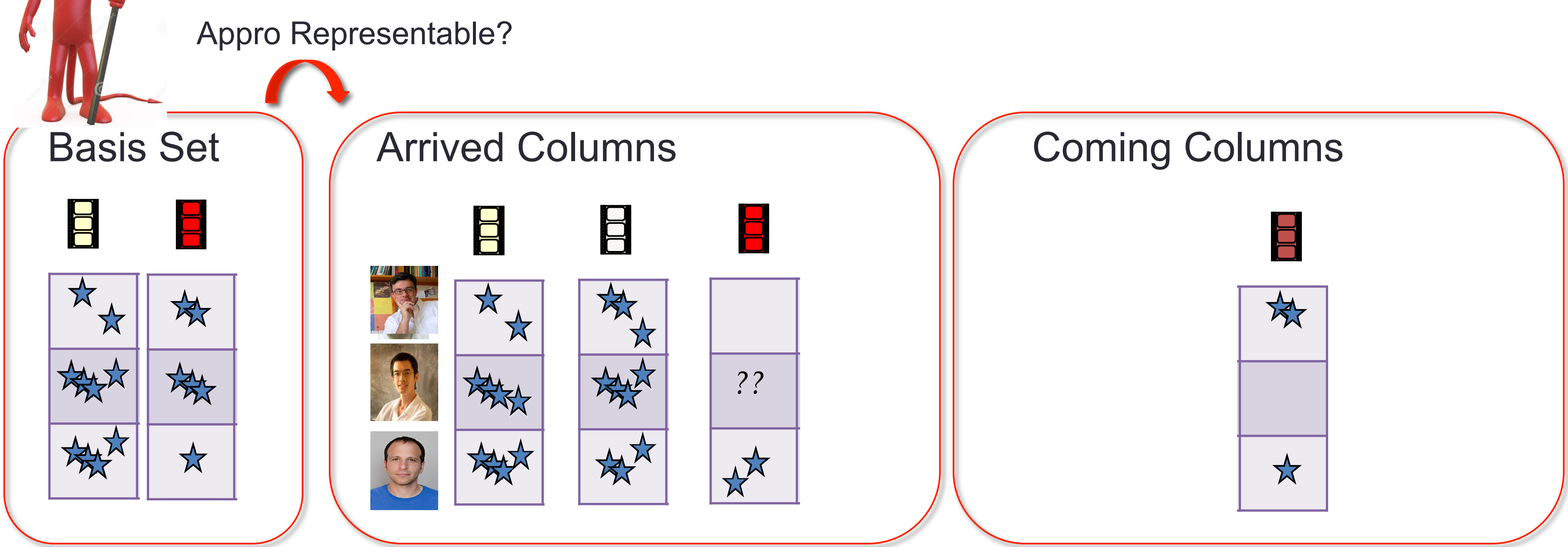
What if the data is noisy?



Krishnamurthy and Singh, Low-Rank Matrix and Tensor Completion via Adaptive Sampling, NIPS 2013

Noisy Life-Long Matrix Completion

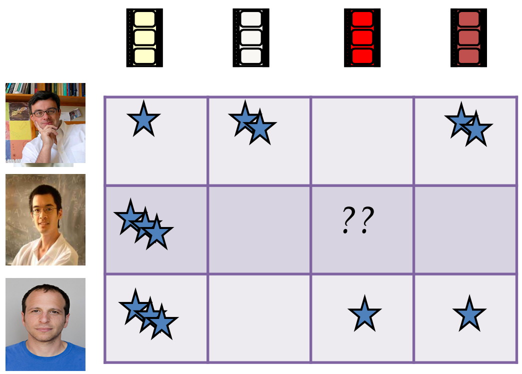
- Challenges in the noisy setting
  - Noise might be adversarial
  - Noise propagates as the data comes along online



Outline

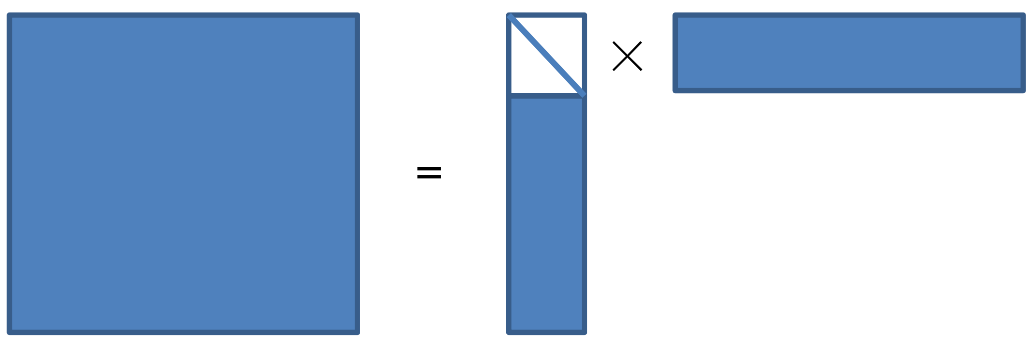
- Motivation and examples
- Our goal and approach
- **Matrix completion background**
- Robustness Analysis
- Experimental Results

Matrix Completion Background



Effectiveness of Low Rankness

- It is a global constraint
- Data compression: *mn* → *r(m+n-r)*
  - Significantly reduces the degrees of freedom: *mn* → *r(m+n−r)*

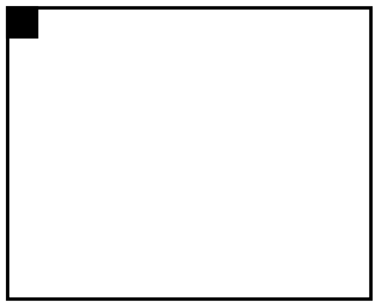


We have to consider the correlation among rows and columns.

Matrix Completion Background (cont'd)

- Incoherence is necessary

X =



$(= e_1 e_1^*)$

- Any subset of entries that misses the (1,1) component tells you nothing!

$X = U\Sigma V^T, \ rank(X) = r$

$\|U^T e_i\|_2 \le \sqrt{\frac{\mu_0 r}{m}}$ *(left incoherence)*

$\|V^T e_i\|_2 \le \sqrt{\frac{\mu_0 r}{n}}$ *(right incoherence)*

X =

$(= e_1 v^*)$

- Still need to see the entire first row

Want each entry to provide nearly the same amount of information.

Related Work

- Matrix completion by nuclear norm.
  - Candès & Tao 2009, Recht 2011, Gross 2011, Zhang et al. 2016
- Matrix completion by alternating minimization
  - Jain et al. 2012, Hardt et al. 2014, Sun & Luo 2015, Ge et al. 2016
- Matrix completion by adaptive sampling
  - Krishnamurthy & Singh, 2013 & 2014

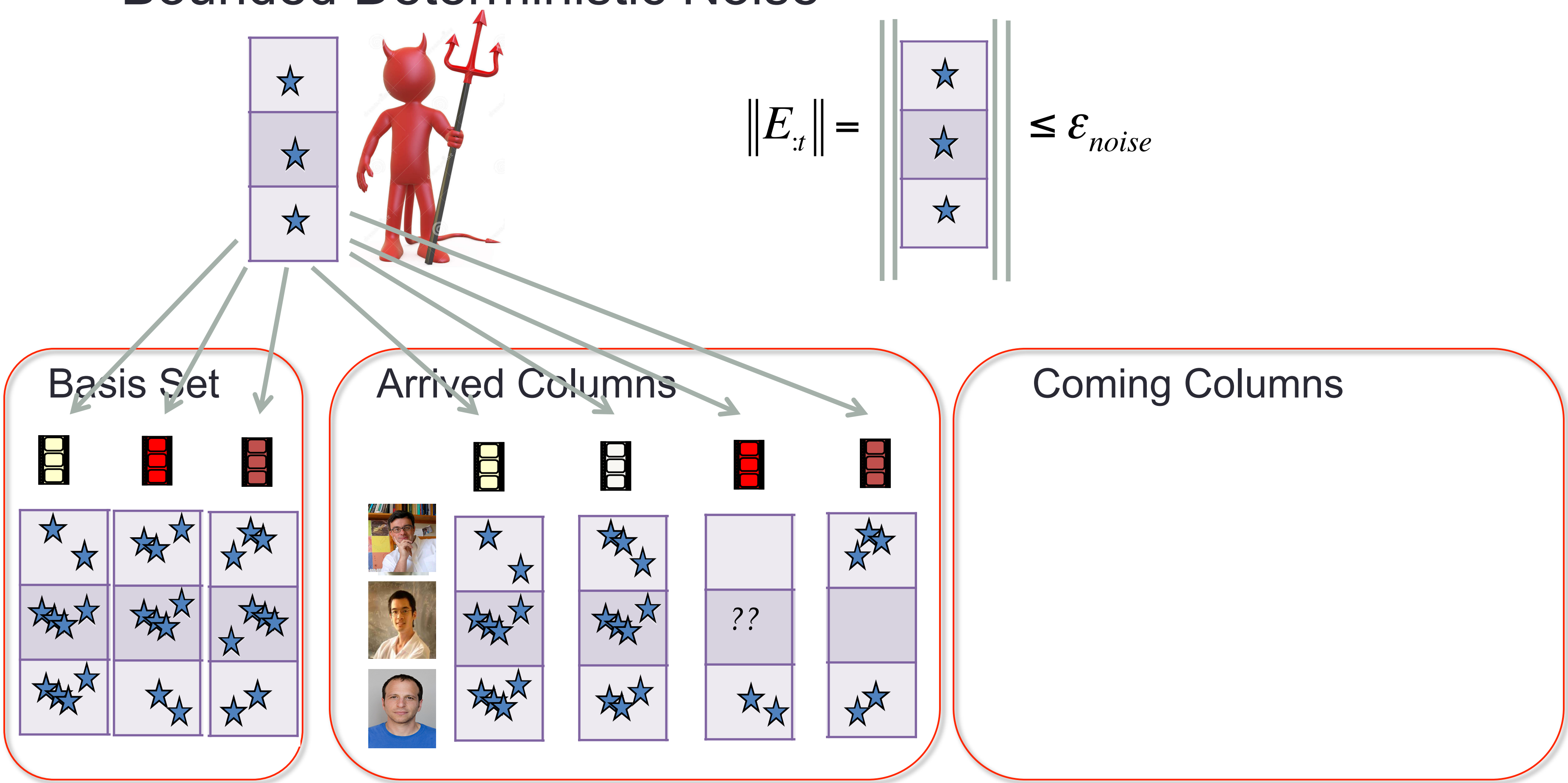
Very little analysis of noise-tolerant online matrix completion algorithm

Outline

- Motivation and examples
- Our goal and approach
- Matrix completion background
- **Robustness Analysis**
- Experimental Results

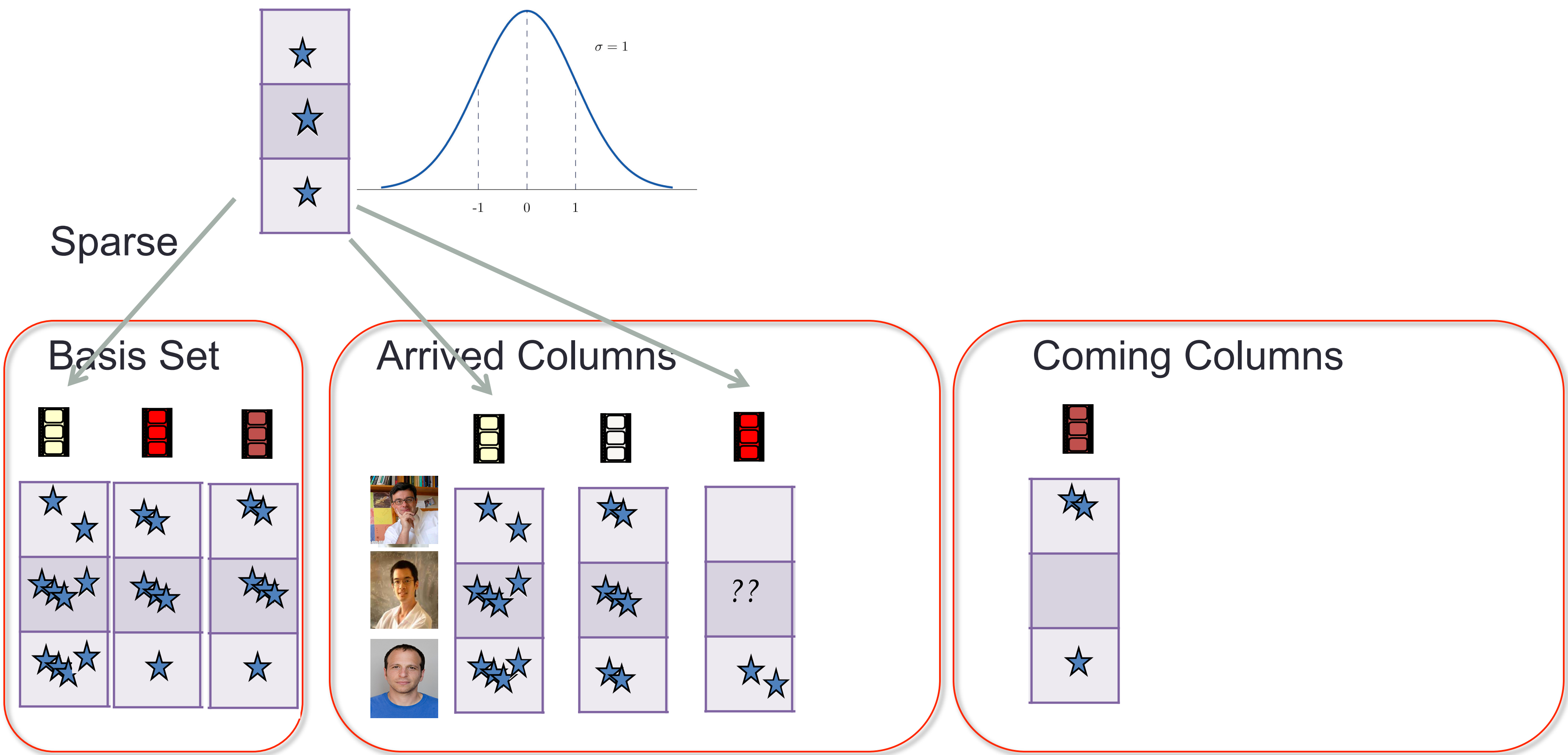
Noise Model

- Bounded Deterministic Noise



Noise Model (cont'd)

- Sparse Random Noise



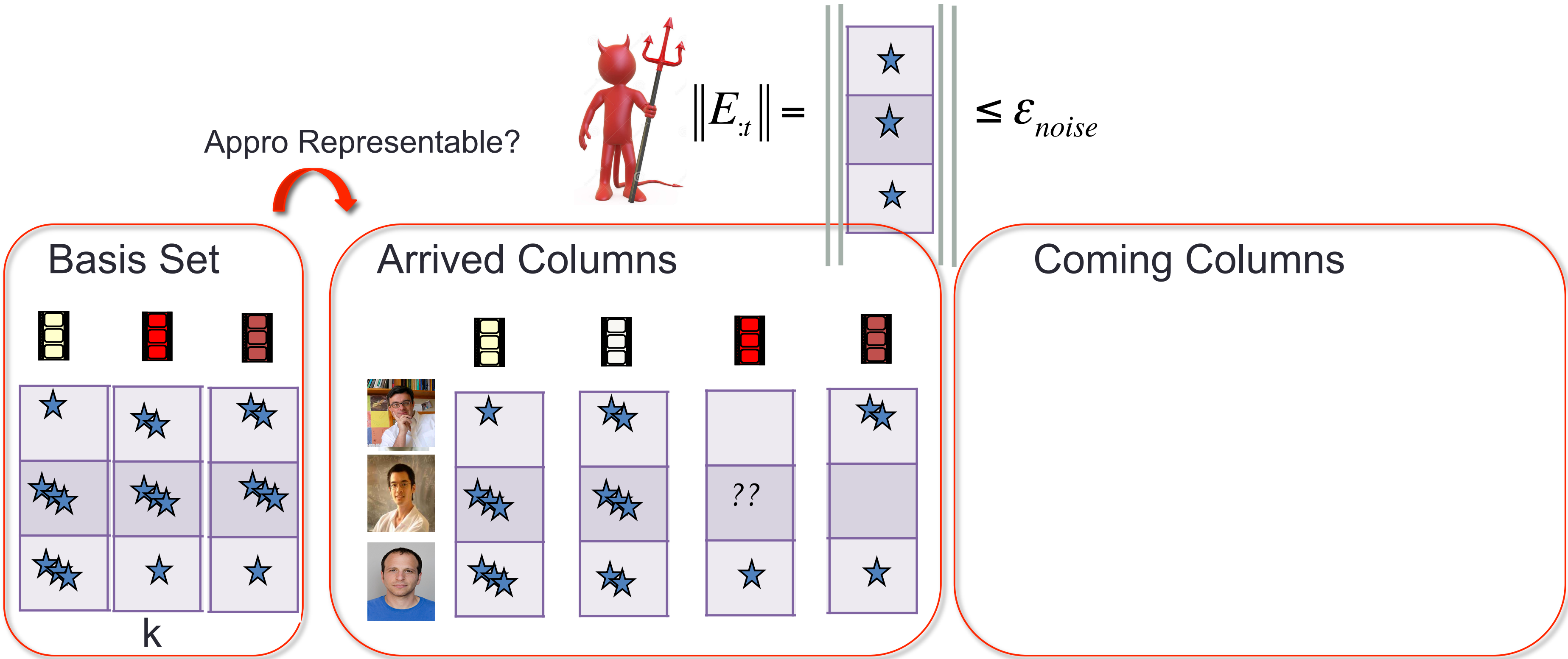
Main Results --- Bounded Deterministic Noise

Theorem (Bounded Deterministic Noise)

Sample Complexity: $O((\mu_0 nr + mnk\epsilon_{noise})\log^2(\frac{r}{\delta}))$

Output Error: $\|\widehat{\mathbf{M}}_{:t} - \mathbf{M}_{:t}\|_2 \leq \Theta\left(\frac{m}{d}\sqrt{k\epsilon_{noise}}\right)$

Parameter: k *number of bases,* ϵ_{noise} *noise magnitude,* r *rank,* μ_0 *incoherence,* δ *failure prob.,* d *unif(d)*



Discussion --- Bounded Deterministic Noise

Theorem (Bounded Deterministic Noise)

Sample Complexity: $O((\mu_0 nr + mnk\epsilon_{noise})\log^2(\frac{r}{\delta}))$

Output Error: $\|\widehat{\mathbf{M}}_{:t} - \mathbf{M}_{:t}\|_2 \leq \Theta\left(\frac{m}{d}\sqrt{k\epsilon_{noise}}\right)$

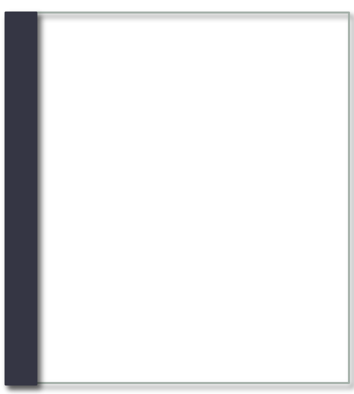
Parameter: k *number of bases,* ϵ_{noise} *noise magnitude,* r *rank,* μ_0 *incoherence,* δ *failure prob.,* d *unif(*d*)*

- The error propagates only in the speed of $\sqrt{k}$, low propagation rate
- Sample Complexity $O(\mu_0 nr\log^2 n)$, if $\varepsilon_{noise} \leq O(\mu_0 r/mk)$
- Incoherence assumption in only one direction

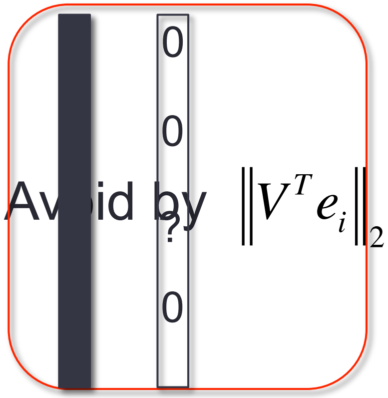
Why left incoherence is enough?

Avoid by $\left\| U^T e_i \right\|_2 \leq \sqrt{\frac{\mu_0 r}{m}}$ (left incoherence)

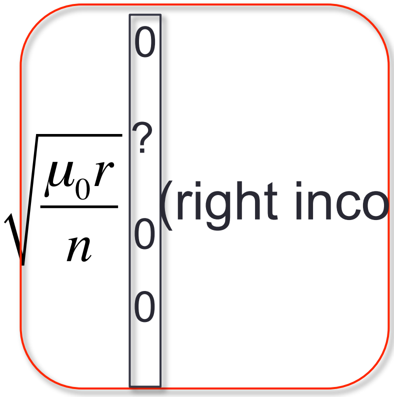
Not an Issue:



Arrived Columns



Coming Columns



Avoid by $\left\| V^T e_i \right\|_2 \leq \sqrt{\frac{\mu_0 r}{n}}$ (right incoherence)

Proof Sketch --- Bounded Deterministic Noise

Fact 1

$$U^k = span\{u_1, u_2, ..., u_k\}$$
$$\tilde{U}^k = span\{\tilde{u}_1, \tilde{u}_2, ..., \tilde{u}_k\}$$

If $\theta(u_i, \tilde{u}_i) \le \varepsilon_{noise}$ and $\theta(\tilde{U}^{i-1}, \tilde{u}_i) \ge \sqrt{20 i \varepsilon_{noise}}$

Then $\theta(U^k, \tilde{U}^k) \le \gamma_k / 2$

Proof idea: Reduction on *k*

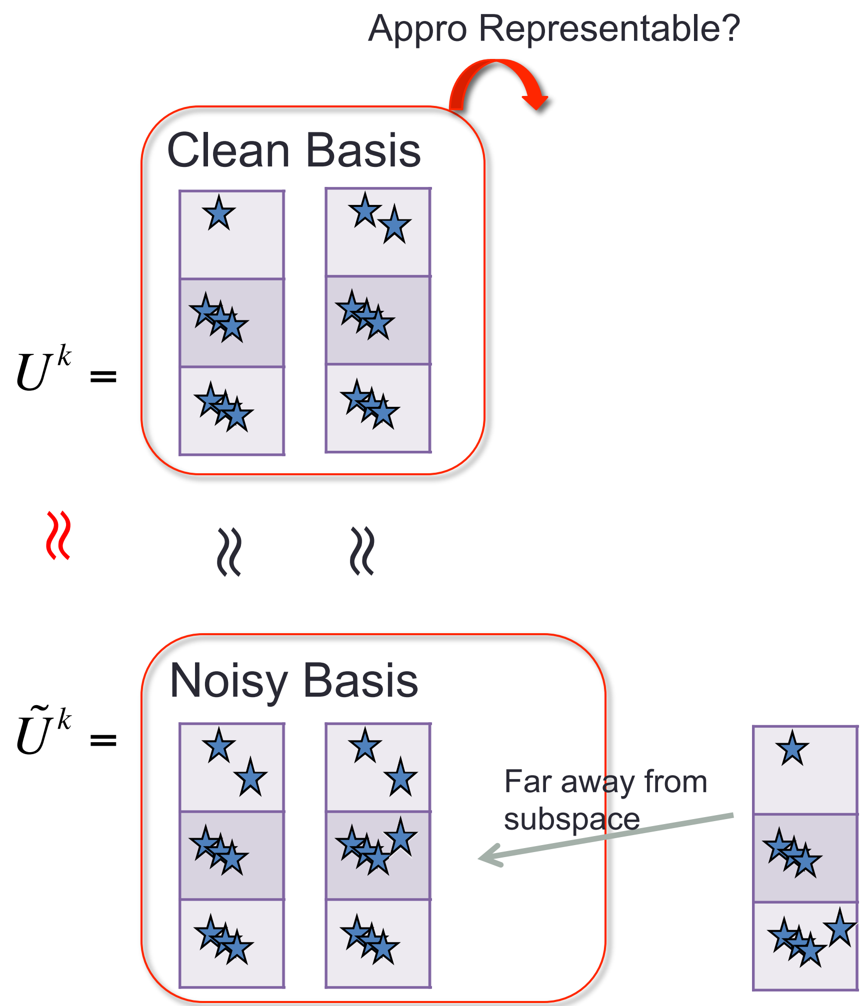
Fact 2 $U^k \approx \tilde{U}^k$

Concentration

$$\theta(M_{:t}, U^k) \approx \theta(M_{:t}, \tilde{U}^k) \qquad \left\| M_{\Omega t} - P_{\tilde{U}^k_{\Omega:}} M_{\Omega t} \right\|_2 \approx \frac{d}{m} \left\| M_{:t} - P_{\tilde{U}^k} M_{:t} \right\|_2$$
$$= f(\theta(M_{:t}, \tilde{U}^k))$$

$\theta(M_{:t}, U^k)$ is determined by $\left\| M_{\Omega t} - P_{\tilde{U}^k_{\Omega:}} M_{\Omega t} \right\|_2$

small



Main Results --- Sparse Random Noise

Theorem (Sparse Random Noise, Upper Bound)

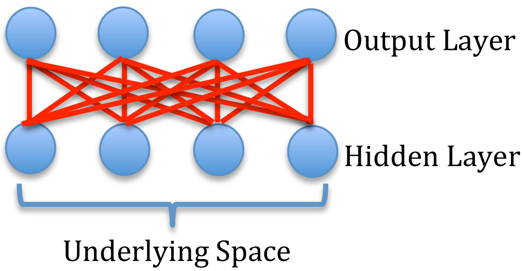
Noise Sparsity: $s_0 \leq O(m)$
Noise Magnitude: Arbitrarily large
Sample Complexity: $O(\mu_0 rn \log\left(\frac{r}{\delta}\right))$
Output Error: Exact Recovery

Theorem (Sparse Random Noise, Lower Bound)

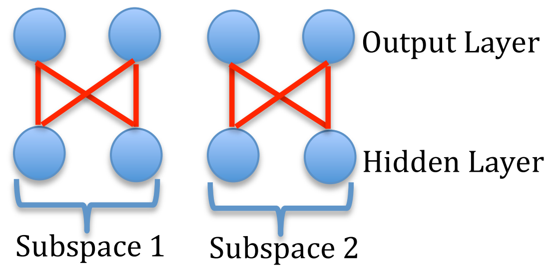
Sample Complexity: $\Omega\left(\mu_0 rn \log\left(\frac{r}{\delta}\right)\right)$
Output Error: Exact Recovery

	Passive Sampling	Adaptive Sampling
Complexity	$\mathcal{O}\left(\mu_0 nr \log^2(n/\delta)\right)$[22]	$\mathcal{O}\left(\mu_0 nr \log^2(r/\delta)\right)$[19] $\mathcal{O}\left(\mu_0 nr \log(r/\delta)\right)$ (Ours)
Lower bound	$\mathcal{O}\left(\mu_0 nr \log(n/\delta)\right)$[10]	$\mathcal{O}\left(\mu_0 nr \log(r/\delta)\right)$ (Ours)

Main Results --- Mixture of Subspaces



(a) Single Subspace



(b) Mixture of Subspaces

Theorem (Mixture of Subspaces, Sparse Random Noise)

Noise Sparsity: $s_0 \leq O(m)$

Noise Magnitude: Arbitrarily large

Sample Complexity: $O(\mu_\tau \tau^2 n \log(\frac{r}{\delta}))$ *(Single:* $O(\mu_0 r n \log(\frac{r}{\delta}))$*)*

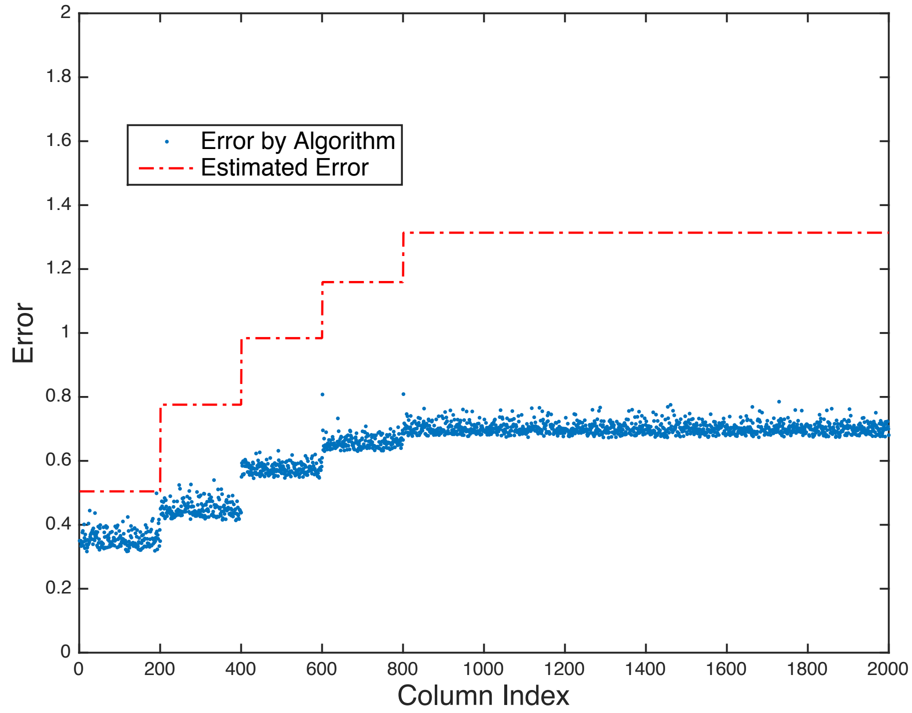
Outpur Error: Exact Recovery

Paramter: μ_τ *incoherence of each subspace,* τ *dimension upper bound of each subspace*

Outline

- Motivation and examples
- Our goal and approach
- Matrix completion background
- Robustness Analysis
- **Experimental Results**

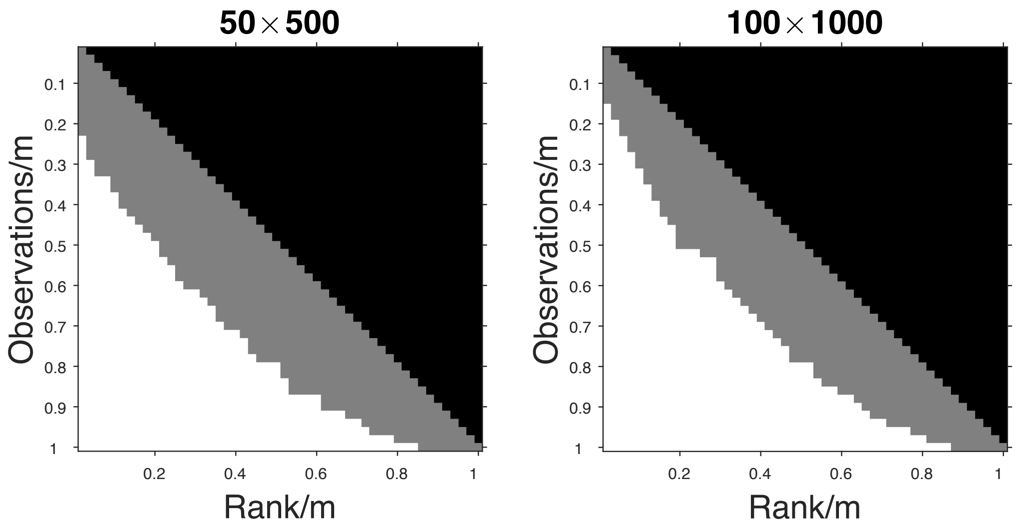
Experiment Results --- Bounded Deterministic Noise



$$\mathbf{L} = \left[\mathbf{u}_1 \mathbf{1}_{200}^T, \sum_{i=1}^{2} \mathbf{u}_i \mathbf{1}_{200}^T, \sum_{i=1}^{3} \mathbf{u}_i \mathbf{1}_{200}^T, \sum_{i=1}^{4} \mathbf{u}_i \mathbf{1}_{200}^T, \sum_{i=1}^{5} \mathbf{u}_i \mathbf{1}_{1,200}^T\right] \in \mathbb{R}^{100 \times 2{,}000}$$

Noise Magnitude: $\epsilon_{noise} = 0.6$

Experiment Results --- Sparse Random Noise



White Region: Nuclear norm minimization succeeds.
White and Gray Regions: Our algorithm succeeds.
Black Region: Our algorithm fails.

Experiment Results --- Mixture of Subspaces

Table 2: Life-long Matrix Completion on the first 5 tasks in Hopkins 155 database.

#Task	Motion Number	$d = 0.8m$	$d = 0.85m$	$d = 0.9m$	$d = 0.95m$
#1	2	9.4×10^{-3}	6.0×10^{-3}	3.4×10^{-3}	2.6×10^{-3}
#2	3	5.9×10^{-3}	4.4×10^{-3}	2.4×10^{-3}	1.9×10^{-3}
#3	2	6.3×10^{-3}	4.8×10^{-3}	2.8×10^{-3}	7.2×10^{-4}
#4	2	7.1×10^{-3}	6.8×10^{-3}	6.1×10^{-3}	1.5×10^{-3}
#5	2	8.7×10^{-3}	5.8×10^{-3}	3.1×10^{-3}	1.2×10^{-3}

Summary

- Life-Long Matrix Completion
  - Online
  - Noise Tolerant
- Sample Complexity
  - Bounded Noise: As small as noiseless case
  - Sparse Noise: Achieve lower bound in the worst case, better than nuclear norm minimization method
  - Mixture of Subspaces: Potential smaller sample complexity

Thank You